Statistical Inference Through Variable Adaptive Threshold Algorithm in Over-Sampling the Imbalanced Data Distribution Problem

S. Karthikeyan & T. Kathirvalavakumar

Abstract

Classification of an imbalanced data is a major problem in many real-time systems. Classifiers classify the majority class samples with lesser misclassification and the minority class samples with more misclassification. The biased decisions of a classifier are due to the less availability of samples in the minority class. To solve this problem, new re-sampling method for over-sampling is proposed. The synthetic samples for the minority class are generated by statistically analysing the features of a minority class samples. Here, the samples are generated in double the number of minority samples to reduce its misclassification. It helps the classifier to have a balanced focus on majority and minority classes. The samples over-sampled through this approach are compared against the over-sampling approaches such as SMOTE and ADASYN. The results obtained through the proposed work show better classification accuracy and reduced misclassification rate in both majority and minority classes. Result is evaluated using the statistical evaluation metrics. It is also observed that over-sampling with the proposed approach is better for the small and medium imbalanced ratio datasets.

Keywords

Imbalanced data   Over-sampling   Classification   Statistical methods   SMOTE   ADASYN